AOMedia Research Symposium - Abstracts

Coding Algorithms

Adaptive Optimal Linear Estimators for Enhanced Motion Compensated Prediction
Kenneth Rose, University of California, Santa Barbara

This talk presents approaches that leverage principles of optimal linear estimation to enhance motion compensated prediction. Specifically, the approach considers nearby motion vectors in the motion field as pointers to multiple noisy observations of a current pixel. An optimal linear estimator is derived and employed to optimally combine information extracted by several relevant motion vectors. The general ideas are applied within two settings of interest. In one setting, the motion field is directly obtained from standard block-based motion estimation performed for the current target frame. Observations, due to several neighboring motion vectors, are combined by an optimal linear estimator whose coefficients are obtained from an appropriately derived model, which incorporates both the Markov properties of video signals and the decay of influence of motion vectors with distance from their location. The linear estimator automatically adapts to local statistics and configurations. The second setting is that of bi-directional prediction, where existing (available to the decoder) motion information relating the reference frames, is utilized to generate a "free" reference frame that is co-located with the current frame. It employs an optimal linear estimator whose coefficients are determined by accounting for the match between pixels pointed to in the respective reference frames, thus achieving both precision and adaptivity to local statistics. The resulting co-located interpolated frame is treated as an additional reference frame, allowing for further offset adjustment as usual via estimated “motion vectors”. The proposed methods were implemented in the AV1 codec and experimental results show significant performance gains.

What Machines Can Learn from Humans About Lossy Compression
Tsachy Weissman, Stanford University

Inspired by Shannon's work on estimating the entropy of a language, we experimented with a framework for image compression comprising one human describing images using text instructions to another, who is tasked with reconstructing the original image. These image reconstructions were then rated by human scorers on the Amazon Mechanical Turk platform and compared to reconstructions obtained by existing image compressors. We've conducted similar experiments for audio compression. The insights gleaned offer a perspective on the potential for substantial improvements over current approaches to lossy compression and related information processing tasks.

A Switchable Region-Based Coding Tool for AV1 Video Codec

Fengqing Maggie Zhu, Purdue University

Significant advances in video coding has been developed in the last two decades to satisfy the growing requirements of video applications. In this talk, we introduce the design and development of a switchable region-based coding tool that leverages semantic segmentation to achieve coding gain and to contribute to the AOMedia project efforts. Current video coding standards utilize hybrid coding techniques consisting of 2D transforms and motion compensation techniques to remove spatial and temporal redundancy. Our approach is different in that we only encode regions of a video frame that are perceptually significant. The perceptually insignificant regions are not encoded. By perceptually insignificant pixels we mean regions in the frame that an observer will not notice any difference without observing the original video sequence, which are typically highly textured regions. The encoder fits a model to the perceptually insignificant pixels in the frame and transmits the model parameters to the decoder as side information. The decoder uses the model to reconstruct the pixels. We propose a semantic segmentation approach using a two-stream cascade network to identify regions to skip encoding. A switchable mechanism is developed to determine when to use the proposed coding mode for each scene. We present a new perceptual quality assessment measure for the region-based coding tool. The proposed method is evaluated on standard test sets and the YouTube UGC datasets, which showed significant data rate reduction with satisfying visual quality.

Incorporating Physical Modeling into Deep Generative Networks for Image and Video Compression

Aswin Sankaranarayanan, Carnegie Mellon University

What is an image? The answer to this question or, equivalently, a precise characterization of the space of images (say, natural images) is the building block for solving many fundamental problems in modern image processing, computer vision and graphics. For example, sparsity of images in DCT and wavelet bases have led to numerous advances in image processing (compression, denoising) and imaging (compressive sensing). Sparsity of image gradients have provided tools for solving many inverse problems in image restoration. Similarly, learning based models based on sparse dictionaries and deep neural networks have provided dramatic improvements in our ability to solve image processing problems. It is not a stretch to suggest that a foundational breakthrough for modeling the space of images, along with associated tools for optimization, will provide the seed for fundamental breakthroughs in all disciplines where images play a central role. This project aims to advance image and video modeling using data-driven generative models and associated tools for advancing compression of the signals.

Coding Efficiency Evaluation of AV1 Coding Tools

Ryan Lei, Intel

AV1 is an emerging royalty-free video coding standard developed by the Alliance of Open Media (AOMedia) industry consortium. The main goal of AV1 is to achieve substantial coding efficiency gain while maintaining practical implementation complexity. It was finalized in 2018 and has gain great attraction from the industry since then. Multiple software

implementations of the AV1 decoder and encoder have been made available, and hardware implementations of the decoder and encoder are also emerging. For any practical implementation of the video encoders based on a video coding standard, it is critical to understand the relative quality/coding efficiency gain and complexity from individual coding tools proposed in the standard. Based on that, decision can be made by encoder vendors on what set of coding tools should be supported in the encoder, which gives optimal trade-off for quality, complexity and cost. In order to evaluate coding efficiency gain from individual coding tools, different approaches can be taken. For example, if a vendor has already implemented encoders based on previous generation of video coding standards and want to support AV1 leveraging existing implementation, then the existing implementation can be used as baseline and individual AV1 coding tools can be added on top of that. In this scenario, coding efficiency gain from each individual tool can be evaluated by enabling it from a low quality baseline. This is typically called “tool on” test. In another scenario, if a vendor want to start the implementation using AOM reference encoder as baseline and disable certain coding tools in order to achieve certain performance target, coding efficiency loss can be evaluated by disabling it from a high quality baseline. This is typically called “tool off” test. In this work, coding efficiency gain for majority of the coding tools proposed in the AV1 coding standard are evaluated using the “tool on” test following the standard test conditions defined by the AOM committee. In order to do that, AV1 reference encoder code is modified to support passing control flags through the config file to turn on/off individual coding tools in the reference encoder. Based on that, extensive encoding tests are executed to check coding efficiency gain of each individual coding tool against the baseline, which disables all new coding tools and use a configuration that is close to VP9 as much as possible. In this study, standard BDRATE metric is used to calculate relative coding gain against the baseline encoder. Average BDRATE is calculated across a whole set of test sequences included in the AV1 “objective-1-fast” test set. In order to eliminate the impact of different bitrate control algorithm, constant QP mode is used in the encoding process. Three encoding configurations are tested, including: All Intra, Low-Delay mode, and High-Delay mode. This paper summarizes the result of the tests and provides a high level ranking of different coding tools in terms of their coding efficiency gain. Effectiveness of few critical coding tools are also further investigated.

An Overview of New Experimental Coding Tools
Sarah Parker, Google

Although the industry is currently focused on the implementation and optimization of AV1, AOMedia Research is continuing to develop new coding tools that deliver higher coding gains within acceptable complexity bounds. At this stage the work is exploratory in nature. This presentation provides a brief technical overview of the new tools currently under development in the experimental branch of the reference codebase.

Performance & Optimization

Evaluating Video Codecs Through Objective and Subjective Assessments
Fan Zhang, University of Bristol

In this work, the performance of state-of-the-art video codecs, High Efficiency Video Coding (HEVC) Test Model (HM), AOMedia Video 1 (AV1) and Versatile Video Coding (VVC) Test Model (VTM), is evaluated through objective and subjective quality assessments. Nine source sequences were carefully selected to offer both diversity and representativeness, and their different resolution versions were encoded by both codecs towards pre-defined target bit rates. The compression efficiency and computational complexity of these codecs were first evaluated through objective quality assessment, and the subjective quality of their compressed content was further tested by conducting psychophysical experiments. Moreover, the Dynamic Optimizer method was also employed to compare HM and AV1 codecs across different resolutions in wider bit rate ranges. All the subjective quality scores collected in these experiments were then employed to evaluate the correlation performance of popular objective video quality metrics. The selected source sequences, compressed video content and associated subjective data have now been made available online, offering a valuable resource for compression performance evaluation and objective video quality assessment.

Speeding up VP9 Intra Encoder with Hierarchical Deep Learning Based Partition Prediction
Somdyuti Paul, University of Texas at Austin

In VP9 video codec, the sizes of blocks are decided during encoding by recursively partitioning 64×64 superblocks using rate-distortion optimization (RDO). This process is computationally intensive because of the combinatorial search space of possible partitions of a superblock. We propose a deep learning based alternative framework to predict the intra-mode superblock partitions in the form of a four-level partition tree, using a hierarchical fully convolutional network (H-FCN). We created a large database of VP9 superblocks and the corresponding partitions to train an H-FCN model, which was subsequently integrated with the VP9 encoder to reduce the intra-mode encoding time. The experimental results establish that our approach speeds up intra-mode encoding by 69.7% on average, at the expense of a 1.71% increase in the Bjontegaard-Delta bitrate (BD-rate). While VP9 provides several built-in speed levels which are designed to provide faster encoding at the expense of decreased rate-distortion performance, we find that our model is able to outperform the fastest recommended speed level of the reference VP9 encoder for the good quality intra encoding configuration, in terms of both speedup and BD-rate.

<Nathan Egge - TBD>

Learning-Based AV1 Optimization for VoD and RTC Use Cases
Jinaa Liu, Visionular

In this presentation, we will present the most recent results we obtained in the optimization of AV1 encoding using our framework namely Aurora, for both use cases of video on demand (VoD) and real-time communications (RTC). We will mainly describe our end-to-end solution for video encoding and decoding leveraging learning based approaches, including content differentiated coding tool deployment, content adaptive preprocessing and postprocessing, neural network based mode decision, etc. The use of machine learning in

our approaches demonstrate a great potential for effective AV1 encoding optimization, confirmed by the extensive experimental results collected by comparing our Aurora encoder against x264, x265, and the open source AV1 codebase libaom, over a variety of video content. Future work will also be addressed. Continuous optimization work on AV1 will lead to its eventual wide deployment for all kinds of scenarios.

Still Picture

AVIF: Overview and Compression Performance
Cyril Concolato, Netflix

Abstract to come.

Applying Video Coding Tools to WebP Images
Pascal Massimino, Google

This talk covers the Google WebP team's experience developing web image formats from video codecs. We cover design decisions based on VP8 / WebP, as well as AV1 - how the underlying intra tools in these video codecs affects users via memory footprint, feature support, etc.

Additionally, the team will talk about ongoing research on a new WebP successor. The goal of these research tools is a (1) lightweight image format capable of software decode on consumer devices that (2) reduces bits on the wire for web images, while including (3) key features like lossless, alpha, animation, and HDR.

ML-Based Encoding

Opportunities to use Neural Media Compression
George Toderici, Google

The primary purpose of this talk is to expose practitioners of compression technologies to neural methods that are used in image/video compression. The hope is that some of these technologies can be translated into the next generation of image/video codecs. I will cover methods ranging from those that are quite experimental to some that have a proven track record, showing that they are capable of winning multiple compression challenges.

Deep Learning for Image Compression
Yao Wang, New York University

We will present our work exploring the use of deep learning for different aspects of image and video compression, including intra prediction using an encoder+denoising approach, video prediction using dynamic deformable convolution, and scalable image compression.

Deep Neural Network Based Frame Reconstruction For Optimized Video Coding - An AV2 Approach

Dandan Ding, Hangzhou Normal University

In this report, I would like to share my ideas and results on CNN-based in-loop and out-loop filtering, aiming to develop new coding tools for the next video coding standard AV2. After investigating various work in in-loop filtering, we propose our CNN model, which is incorporated into AV1 for both intra and inter coding. We have designed various CNN structures and it is interesting to investigate which kind of CNN structures is more effective for handling loop filtering problem. Besides, I will share our recent results on multi-frame video enhancement.

Deep Video Precoding
Yiannis Andreopoulos, iSIZE Technologies

A core technical objective of iSize is to find the optimal way to preprocess (or precode – in our nomenclature) any input video into a (typically) smaller pixel stream, in order for advanced video encoders like AOMedia VP9 and AV1 to achieve the best video quality at a given bitrate, or save the maximum amount of bits for a given perceptual quality level. We are especially interested in this problem given that perceptual quality can now be measured with advanced perceptual quality metrics from the literature, e.g., using the fusion of multiple quality metrics (akin to VMAF of Netflix) or via other advanced metrics like DeepQA.

In this talk, we shall summarize the concepts behind our approach and show some promising results with our existing framework that is available for testing or usage at bitsave.tech. We shall also review recent viewer preference tests done via the Amazon Mechanical Turk service, where independent MTurk workers from around the world viewed full HD video on their devices and selected their preference between the one encoded with and without the use of our deep precoding engine.

Perceptual Metrics

Perceptually Optimizing Deep Image Compression
Li-Heng Chen, University of Texas at Austin

The use of Lp (p=1,2) norms has largely dominated the measurement of loss in neural networks due to their simplicity and analytical properties. However, when used to assess the loss of visual information, these simple norms are not very consistent with human perception. Here, we describe a different "proximal" approach to optimize image analysis networks against quantitative perceptual models. Specifically, we construct a proxy network, broadly termed ProxIQA, which mimics the perceptual model while serving as a loss layer of the network. We experimentally demonstrate how this optimization framework can be applied to train an end-to-end optimized image compression network. By building on top of an existing deep image compression model, we are able to demonstrate a 20% average bitrate reduction over MSE optimization, given a specified perceptual quality (VMAF) level.

On Perceptual Coding: Quality, Content Features and Complexity
Patrick Le Callet, University of Nantes

The influence of content characteristics on the efficiency of redundancy and irrelevance reduction in video coding is well known. Each new standard in video coding includes additional coding tools that potentially increase the complexity of the encoding process in order to gain further rate-distortion efficiency. Balancing rate-distortion against computational complexity is crucial these days. In this talk, I will highlight several key questions and possible avenues to address them with some flavors of perception and cognitive computing.

Physical Modeling

Informing Video Compression With Physical Simulation
Theodore Kim, Yale University

Video coding and realistic physical simulation share a common goal: compactly predicting the state of the world at the next moment in time. Physical simulation algorithms have been developed to predict the visual motion of objects over time, and have been successful in generating content for movies, television, and games. But, can they be used to inform video codecs? Prediction using direct numerical simulation is too computationally expensive, but fast, approximate simulations have been developed that may provide insight. In particular, "subspace" methods excel at efficiently "playing back" existing scenes in a way that is already reminiscent of video decoding, and use familiar methods such as DCT compression to improve efficiency. In this talk, I will discuss these techniques and their potential applications.

General Compression

Mode-dependent Data-driven Transforms for AV1
Antonio Ortega, University of Southern California

In the AV1 codec multiple transform kernels are available, and the encoder chooses the best transform per block with a rate-distortion search. In this work, we aim to obtain novel data-driven transforms that are tailored for AV1. We adopt a rate-distortion optimized transform (RDOT) learning scheme in order to obtain useful additional transforms for AV1. In particular, the learned RDOTs are mode-dependent: one particular RDOT is trained for each intra mode, to better capture different block characteristics for different modes. As a result, in addition to original transform kernels, three separable RDOTs and one non-separable are included for intra blocks. In addition, eight separable RDOTs are included in inter blocks. With an increased encoder complexity, we achieve around 0.8% gain in bitrate overall, with more than 1% gain on key frames.

MeasuringVideo Quality With VMAF: Why You Should Care
Christos Bampis

VMAF (Video Multi-Assessment Fusion) is a quality metric that combines human vision modeling with machine learning. It demonstrates high correlation to human perception and

gives a score that is consistent across content. VMAF has been widely applied at Netflix in areas such as video quality monitoring and encoding optimization. VMAF was released on Github in 2016 and has had considerable updates since that time. In this talk, we will give a brief review of the history behind VMAF development and present details behind the actual VMAF algorithm. In addition, we will discuss some major adoption cases for VMAF and present a number of challenges that lie ahead.

Motion Based Video Frame Interpolation
Anil Kokaram, Trinity College Dublin

Motion based frame interpolation has been studied in the Digital Video Processing community for many years since the early days of television standards conversion. It is arguably one of the main components of classic hybrid video codecs. Recent exciting developments in using CNNs for interpolation have been achieved without the benefit of knowledge from that television and film community. We compare industrial strength retimers (widely used in the film effects community) with two of the more recent CNN algorithms for frame rate conversion. Results show that CNNs compare favourably with the existing technology but still have some way to grow. We use these observations to show how important it remains to consider motion smoothness explicitly in these kinds of problems and what that means for the future.

AV1 Implementers Forum

Real-Time AV1 with SVC support in WebRTC
Alexandre Gouaillard, CoSMo

While a lot are waiting for AV1 to use it in file-based protocols like HLS and MPEG-DASH (with or without CMAF), a happy few are aiming at using it for Real-Time communications. Cisco demonstrated a first (proprietary) real-time AV1 software encoder fully integrated in WebRTC and then in turn in their video conference system Webex without SVC support, and CoSMo presented another real-time software encoder in WebRTC for streaming in MilliCast based on Open-Source libaom's real-time mode. Since then, CoSMo added support for SVC. Details of implementation of RTP payload, with SVC support, in WebRTC, and multiple benchmark at the protocol and media transport level will be provided.

AV1 in the MilliCast Real-Time (>200ms) Streaming Platform: The System Level Point of View
Richard Blakely, MilliCast

MilliCast leverage CoSMo's "LUXON" media infrastructure composed of client-side implementation of libwebrtc, and Medooze Media serves, among others. The recent addition of software real-time AV1 encoder capacity for broadcasters impacts many aspects of the Streaming platform. How to support real-time AV1 while keeping real-time end-to-end encrypted recording, real-time forensic watermarking, server-side ad-insertion, ... This presentation will list several practical details about what parts of a streaming infrastructure gets impacted by adding AV1, and how one can address them without compromising on the

latency, quality or scale. The server-side ad-insertion into AV1 WebRTC stream leverages AOMedia's INTEL Scalable Codec Technology and would have been presented separately at IBC on September 13th, within the scope of INTEL's Visual Cloud Conference.

SVT-AV1 Encoder
Nader Mahdi, Intel

Scalable Video Technology (SVT) is a standard-agnostic architecture that allows software encoders to scale efficiently using a multi-core Xeon CPU, the main processing unit used in cloud environments. To address the very-high complexity of AV1 encoding, Intel and Netflix have recently partnered to develop SVT-AV1, an efficient & scalable implementation of the AV1 standard, to the open source community under a permissive BSD+Patent license, and they are currently working on completing the implementation of the features, while targeting both the VOD and Live applications. SVT-AV1 has evolved rapidly since it was open sourced, and it is now capable of achieving AOM-AV1 (libaom) levels of quality through both one-pass & two-pass encoding, as well as real-time 1080/4K encoding while maintaining high levels of video quality and offering great bandwidth reduction advantages relative to both AVC and HEVC encoders. In this presentation, we will discuss briefly the SVT architecture, review the supported AV1 features, and present the latest performance-quality results. We will also present the short-term and long-term roadmaps for SVT-AV1, and discuss some of the outstanding challenges.

High-Efficiency AV1 Compression Using dAV1d and Eve
Ronald S. Bultje, Two Orioles

We'll give an overview of the current status and performance (quality/speed/threading) of Videolan's efficient and open-source AV1 decoder, dav1d, and Two Orioles' commercial AV1 encoder, Eve-AV1. We'll aim to point out particular pain points in AV1 in terms of suboptimal performance or complexity, and how these might be improved to make a potential next-generation codec even better.